EP3260: Machine Learning Over Networks
Homework Assignment 1
Due Date: February 7, 2023

Problem 1.1

A differentiable function f is μ-strongly convex iff $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \mu > 0$

$$f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T(\mathbf{x}_2 - \mathbf{x}_1) + \frac{\mu}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2. \tag{1}$$

Then, prove that:

- (1) is equivalent to a minimum positive curvature $\nabla^2 f(\mathbf{x}) \geq \mu \mathbf{I}_d, \forall \mathbf{x} \in \mathcal{X}$;
- (1) is equivalent to $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))^T (\mathbf{x}_2 - \mathbf{x}_1) \geq \mu\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2$;
- (1) implies

  (a) $f(\mathbf{x}) - f^\star \leq \frac{1}{2\mu}\|\nabla f(\mathbf{x})\|_2^2, \forall \mathbf{x}$;

  (b) $\|\mathbf{x}_2 - \mathbf{x}_1\|_2 \leq \frac{1}{\mu}\|\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\|_2, \forall \mathbf{x}_1, \mathbf{x}_2$;

  (c) $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))^T (\mathbf{x}_2 - \mathbf{x}_1) \leq \frac{1}{\mu}\|\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\|_2^2, \forall \mathbf{x}_1, \mathbf{x}_2$;

  (d) $f(\mathbf{x}) + r(\mathbf{x})$ is strongly convex for any convex f and strongly convex r.

Problem 1.2

A function $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth iff it is differentiable and its gradient is L-Lipschitz-continuous (usually w.r.t. norm-2):

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d, \|\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\|_2 \leq L\|\mathbf{x}_2 - \mathbf{x}_1\|_2. \tag{2}$$

For all $\mathbf{x}_1, \mathbf{x}_2$, prove that (2) implies

(a) $f(\mathbf{x}_2) \leq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T(\mathbf{x}_2 - \mathbf{x}_1) + \frac{L}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2$;

(b) $f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T(\mathbf{x}_2 - \mathbf{x}_1) + \frac{1}{2L}\|\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\|_2^2$;

(c) $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))^T (\mathbf{x}_2 - \mathbf{x}_1) \geq \frac{1}{L}\|\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\|_2^2$.

Assume convexity if needed.

Problem 1.3

Define, discuss the benefits, and give examples for the different convergence rates of a sequence of updates $\{\mathbf{x}_k\}$:

(a) Sublinear

(b) Linear

(c) Superlinear

(d) Quadratic

Problem 1.4

Consider

$$\begin{aligned} \text{minimize} \quad & \frac{1}{N} \sum\nolimits_{i \in [N]} f_i(x_i) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

for $\mathbf{b} \in \mathbb{R}^{p \times N}$ and $\mathbf{x} = [x_1, \ldots, x_N]^T$.

(a) Assume strong-convexity and smoothness on f. How would you solve this problem when $N = 1000$?

(b) What if $N = 10^9$?

(c) Can we use Newton's method for $N = 10^9$? Try efficient method for computing $\nabla^2 f(\mathbf{x}_k)$ for $p = 1$ and $b = 1$ (probability simplex constraint). Extend it to $1 \le p \ll N$.

(d) Now, add twice differentiable $r(\mathbf{x})$ to the objective and solve (a)-(c).

Problem 1.5

In the convergence proof of GD with constant step size and strongly convex objective function (see slides), prove the coercivity of the gradient:

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \ge \frac{\mu L}{\mu + L} \|\mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{\mu + L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2.$$